LETTER **Communicated by Daniel Durstewitz**

A Computational Model of the Functional Role of the Ventral-Striatal D2 Receptor in the Expression of Previously Acquired Behaviors

Andrew James Smith
andys@mcmaster.ca
Suzanna Becker
becker@mcmaster.ca
Psychology Department, McMaster University, Hamilton, Ontario, Canada L8S 4K1

Shitij Kapur
Shitij_Kapur@camh.net
Center for Addiction and Mental Health, Toronto, Ontario, Canada, M5R 1T8

The functional role of dopamine has attracted a great deal of interest ever since it was empirically discovered that dopamine-blocking drugs could be used to treat psychosis. Specifically, the D2 receptor and its expression in the ventral striatum have emerged as pivotal in our understanding of the complex role of the neuromodulator in schizophrenia, reward, and motivation. Our departure from the ubiquitous temporal difference (TD) model of dopamine neuron firing allows us to account for a range of experimental evidence suggesting that ventral striatal dopamine D2 receptor manipulation selectively modulates motivated behavior for distal versus proximal outcomes. Whether an internal model or the TD approach (or a mixture) is better suited to a comprehensive exposition of tonic and phasic dopamine will have important implications for our understanding of reward, motivation, schizophrenia, and impulsivity. We also use the model to help unite some of the leading cognitive hypotheses of dopamine function under a computational umbrella. We have used the model ourselves to stimulate and focus new rounds of experimental research.

1 Introduction

Dopamine is a neuromodulator of great interest because of its central role in reward and motivation, as well as in a number of human disorders, including schizophrenia, attention deficit hyperactivity disorder (ADHD), drug addiction, and Parkinson's disease. The precise role of dopamine in each of these processes is still a matter of debate, and a number of partially overlapping hypotheses exist. With a view to formalizing and uniting some of these hypotheses, we present a novel computational model of dopamine function that offers a consistent account of some apparently diverse results

from the animal behavior literature. In particular, we concentrate on the effect of dopamine manipulation on the expression of previously acquired behaviors.

There are several competing hypotheses of dopamine function, which we briefly review. Some are computational and others cognitive, with the computational approaches tending to be oriented around the phasic dopamine response and its role in learning, and the cognitive hypotheses based on behavioral data and possibly pertaining more to the tonic or constant background signal.

The anhedonia hypothesis (Wise, 1982; Wise, Spindler, DeWit, & Gerber 1978) suggested that dopamine mediates the hedonia associated with rewarding environmental stimuli. Evidence was drawn from animal experiments in which neuroleptics (dopamine-blocking drugs) caused extinction-like effects. Extinction refers to the slow disappearance of a conditioned behavior if the resulting reward is not forthcoming. Recent data have eroded the anhedonia hypothesis and supported a more subtle role for dopamine in reward, motivation, and salience. For example, Berridge and Robinson (2003) suggest that reward is a multidimensional construct that can be decomposed into (among others) liking (hedonic) and wanting (motivational) components and that dopamine selectively mediates the latter. This position has gained widespread recognition as the incentive salience hypothesis. Evidence is derived from an extreme experiment in which rats that are deprived of almost all dopamine in the ventral and neostriatum simply stop eating, even though they apparently maintain the motoric capability to do so and even though their life depends on the food that is right under their noses! An analysis of affective responses when artificially fed reveals that the hedonic impact of the food ("liking") is apparently unaffected (Berridge & Robinson, 1998). The incentive salience hypothesis is not computational in nature, although McClure, Daw, and Montague (2003) have developed a computational instantiation to account for some basic experimental data.

Salamone, Cousins, and Snyder (1997) suggest that "dopamine in the nucleus accumbens [part of the ventral striatum] is important for responding to conditioned stimuli and . . . to stimuli that are spatially and temporally distant from the organism" (p. 353). General support for the claim that dopamine is implicated in responding for rewards (or avoiding punishments) that are in some way distal from the organism is provided by a wide range of experiments pertaining to conditioned avoidance (Anisman, Irwin, Zacharko, & Tombaugh 1982; Beninger & Hahn, 1983; Blackburn & Phillips, 1989; Courvoisier, 1956a; Grilly, Johnson, Minardo, Jacoby, & LaRiccia, 1984; van der Heyden & Bradford, 1988; Maffii, 1959; Wadenberg, Soliman, Vanderspek, & Kapur, 2001; Stark, Bischof, & Scheich, 1999; Wilkinson et al., 1998), animal (Richards, Sabol, & de Wit, 1999; Wade, de Wit, & Richards, 2000; Cousins, Atherton, Turner, & Salamone, 1996; Salamone et al., 1991; Salamone, Cousins, & Bucher, 1994) and human (de Wit, Enggasser, & Richards, 2002) models of impulsivity, aphagia (Berridge &

Robinson, 1998), and instrumental responding for food (Dickinson, Smith, & Mirenowicz, 2000; Evenden & Robbins, 1983; Fowler, LaCerra, & Ettenberg, 1986; Rolls et al., 1974; Wise & Schwartz, 1981; Wise et al., 1978), drugs (see Wise, 2002, for a review), conditioned reinforcers (Taylor & Robbins, 1984), and electrical brain stimulation (Ettenberg, 1989; Fibiger, Carter, & Phillips, 1976; Rolls et al., 1974; Salamone, Kurth, McCullough, Sokolowski, & Cousins, 1993). Those studies that directly target the ventral striatum suggest that this is the site where this particular effect of dopamine manipulation is occurring (Berridge & Robinson, 1998; Cardinal, Pennicott, Sugathapala, Robbins, & Everitt, 2001; Salamone et al., 1991, 1993, 1994; Salamone, Wisniecki, Carlson, & Correa, 2001).

In contrast to these motivational perspectives, Horvitz (2002) suggests a more attentional role for dopamine in the gating of sensory, reward, and motor processes. Also within this attentional category, Redgrave, Prescott, and Gurney (1999) propose that dopamine plays a role in switching between different behaviors. For example, Phillips, Stuber, Helen, Wrightman, and Carelli (2003) report that rats could be made to stop what they were doing and press a lever for cocaine (a preconditioned behavior) simply by artificially stimulating dopamine release.

From a computational perspective, a highly influential model is the prediction error hypothesis of dopamine function (Hollerman & Schultz, 1998; Schultz, Dayan, & Montague, 1997; Montague, Dayan, & Sejnowski, 1996; Houk, Adams, & Barto, 1995). Electrophysiological recordings from the primate midbrain suggest that the dopaminergic signal is somewhat analogous to the prediction error signal used to drive learning in the temporal difference (TD) learning algorithm (for TD, see Sutton, 1988; Sutton & Barto, 1998). Apart from being able to account for a range of data pertaining to the phasic firing of dopamine neurons, one of the strengths of this hypothesis is that it posits both a cause (error in predicted reward) and effect (update of reward prediction) of dopamine neuron firing within a concrete computational framework. However, a number of criticisms have been discussed (Berridge & Robinson, 1998; Horvitz, 2000, 2002; Redgrave et al., 1999).

One problem for the prediction error hypothesis is that dopamine is released not just following rewarding stimuli and stimuli that predict reward, but also following novel stimuli, aversive stimuli, and even stimuli that predict aversive events (for reviews, see Ikemoto & Panksepp, 1999; Horvitz, 2000, 2002; Salamone et al., 1997; Joseph, Datla, & Young, 2003). Another problem is that dopamine neurons produce not only intermittent burst or phasic firing, but also a constant background tonic firing. Recently Fiorillo, Tobler, and Schultz (2003) have also found an intermediate sustained firing between a stimulus and a subsequent reward. A more general role for dopamine is therefore suggested.

However, perhaps the major consideration is that the prediction error hypothesis primarily posits a role for dopamine in learning (i.e., the first derivative of behavior), and yet there is compelling evidence that dopamine

is also required for the expression (''zeroth derivative'') of previously acquired behaviors (Cousins et al., 1996; Rolls et al., 1974; Maffii, 1959; Berridge & Robinson, 1998; Wade et al., 2000; Richards et al., 1999; Wadenberg et al., 2001). Moreover, following dopamine manipulation, behaviors appear to be affected differently depending on their relationship to the rewarding (or punishing) outcome. Although Montague, Dayan, Person, and Sejnowski (1995), Montague et al., (1996), and McClure et al. (2003) have extended the TD model to include a role for dopamine in biasing action selection, this approach has not yet been used to account for the selectivity of dopamine manipulation on distal versus proximal rewards. It is this selectivity that is the focus of this letter.

A brief note on the motivation of this work is now offered. One of our long-term goals is a better computational understanding of schizophrenia, and in particular psychosis. Since the ventral striatum and the dopamine D2-receptor subtype have been strongly implicated in the disorder, the first step was to look at behavioral studies that pertain to either the ventral striatum or the specific D2-receptor manipulation, or, where possible, both. The aim of this work is not to model the striatal D2 receptor at an anatomical or physiological level, but rather to induce its functional significance at the behavioral level based on the studies referred to above. Since TD methods provide the currently preeminent computational account of phasic dopamine, our first inclination was to adopt and adapt this approach. However, we were unsuccessful in matching the data to TD and will therefore outline an alternative internal model account of motivated behavior. That is not to say that TD cannot address the data considered below, but rather that we were unable to use it to do so in a parsimonious fashion. The internal model approach described below and TD use very different representational techniques with far-reaching implications, and it is important to know which forms the sounder basis for understanding schizophrenia. Behavioral data provide our current constraint, but future work must draw on additional constraints (including physiological and electrophysiological considerations) to resolve the issue to the satisfaction of behavioral, psychiatric, and computational communities.

2 Schizophrenia, Psychosis, and Conditioned Avoidance

Schizophrenia occurs with a global incidence of around 1% and is considered one of the most debilitating human disorders (Kandel, Schwartz, & Jessell, 1991). The symptoms are broadly divided into two categories: the positive and the negative. The negative symptoms, characterized by a lack of motivation, flat affect, anhedonia, eccentric behavior, social isolation, poverty of speech, and a poor attention span, are chronic and effectively untreatable by pharmacological intervention. In contrast, the positive symptoms or psychosis, which consist of delusions, disordered thoughts,

and hallucinations, often occur in acute phases and may be mitigated or prevented with antipsychotic drugs (APDs). Although psychosis is a major component of schizophrenia, it may also occur in nonschizophrenic individuals. For insight into the nature of delusions and hallucinations, see Maher and Ross (1984) and Beck and Rector (2003), respectively.

All current APDs block dopamine; moreover, there is a striking correlation between the ability of these drugs to block the dopamine D2 receptor and the dose required to mitigate psychosis (Kandel et al., 1991). Interestingly, all drugs of abuse cause an increase in dopamine in the nucleus accumbens (for reviews, see DiChiara, 1999; Kauer, 2003; Ikemoto & Panksepp, 1999), and some of these (particularly cocaine and amphetamine) can also induce psychotic symptoms in users (Bell, 1973; Connell, 1958). The number of D2 receptors is increased in the striatum of (unmedicated) schizophrenia patients in postmortem examination, an effect that is particularly pronounced in patients with positive symptoms (Kandel et al., 1991, chap. 55). The nucleus accumbens in particular may be important as a convergence site for a number of brain regions that are implicated in schizophrenia, including PFC, amygdala, and hippocampus (Grace, 2000), not to mention dopamine projections of the VTA. (See also Grace, 1991, for discussion.) These, along with other data, have led to the hypothesis that psychosis is mediated, if not caused, by an excess of dopamine in the limbic system, of which the ventral striatum is a key component (Kandel et al., 1991, chap. 55).

An important preclinical drug test for potential antipsychotic efficacy is a well-established animal experimental paradigm called *conditioned avoidance* (CA) (Kilts, 2001; Arnt, 1982; Janssen, Niemegeers, & Schellekens, 1965; Wadenberg & Hicks, 1999). The standard CA experiment finds that if a neutral stimulus, such as an auditory tone (the conditioned stimulus or CS), regularly precedes an electric shock (unconditioned stimulus, or US), an animal will learn to avoid the shock by taking appropriate evasive action in response to the tone (Kamin, 1954; Low & Low, 1962; Black, 1963). Appropriate evasive action often involves the animal's running or jumping to another compartment in the cage. An avoidance response is recorded if the animal runs during the tone, and an escape response is recorded if the animal waits until the arrival of the shock. A failure is recorded if the animal fails to run even when shocked.

It is well established that low (noncataleptic) doses of all APDs, administered after the avoidance behavior has been acquired, selectively disrupt that avoidance response yet leave the escape response intact (Ader & Clink, 1957; Arnt, 1982; Cook & Weidley, 1957; Cook & Catania, 1964; Courvoisier, 1956b; Davidson & Weidley, 1976; Ponsluns, 1962). As the drug wears off, the avoidance response is restored (Wadenberg et al., 2001; Smith, Li, Becker, & Kapur, 2004). It seems unlikely that these effects are due to the interaction of the drug with the learning processes of the animal because of the differences in the timescales involved. For example, a rat typically requires many trials over many days to acquire the avoidance response. If an APD is

then administered and the rat is tested 20 minutes later (allowing the drug to take effect), the rat will stop running in response to the tone almost immediately. Similarly, when the rat is tested the following day drug free, the avoidance response will reappear almost immediately. Also, in untreated rats, the extinction of the avoidance response when the shock is actually discontinued tends to be a relatively slow process (Kamin, 1954). Therefore, the immediate impact of the drug is again striking. However APDs do also retard acquisition of the conditioned response, and dissociating the role of dopamine in performance and learning is not straightforward (Beninger, 1989).

The degree of APD-induced avoidance disruption has been correlated with D2-receptor blockade, leading to the suggestion that blockade of this dopamine receptor is the neurochemical link between conditioned avoidance disruption in rats and antipsychotic action in people (Wadenberg et al., 2001). Importantly, conditioned avoidance can also be disrupted by direct intra-accumbens injections of D2-receptor antagonists (blocking D2 receptors) and by accumbens 6-OHDA lesions, destroying the dopamine-releasing neurons themselves (see Ikemoto & Panksepp, 1999, for a review). However, the common behavioral or psychological processes are currently unknown. For example, existing hypotheses of why APDs disrupt avoidance include the inhibition of an internal "fear" or "anxiety" (Cook & Weidley, 1957; Miller, Murphy, & Mirsky, 1957; Davis, Capehart, & Llewellin, 1961; Hunt, 1956), motor impairment (Ponsluns, 1962; Cook & Catania, 1964; Morpurgo, 1965; Beninger, Mason, Phillips, & Fibiger, 1980a, 1980b; Grilly et al., 1984; Aguilar, Mari-Sanmillan, Mortant-Deusa, & Minarro, 2000; Ogren & Archer, 1994), reduced responsiveness to external stimuli (Dews & Morse, 1961), decrease in sensory stimulation (Irwin, 1958; Key, 1961), and loss of attention or arousal (Low, Eliasson, & Kornetsky, 1966). However, following the incentive salience hypothesis of dopamine function, we will propose a motivational explanation that can subsequently be used to account for additional data from other experimental paradigms.

3 The Model

Our model is based on the assumption that an animal builds an explicit internal model of its environment. Internal models have played a significant role in the methodologies of a number of fields, including AI. Indeed, they have been used not only within formal reinforcement learning (reviewed in Sutton & Barto, 1998), but also to model animal conditioning (Schmajuk, 1988) and the dopamine system (Schmajuk, Cox, & Gray, 2001; Suri, Bargas, & Arbib, 2001; Suri, 2001). Although the types of representation used by animals are many and varied (Balleine, Garner, Gonzalez, &

Dickinson, 1995; Berridge & Robinson, 1998; Cardinal, Parkinson, Hall, & Everitt, 2002; Dickinson, 1980), the evidence that animals internally represent action-outcome relationships (among others) is compelling (Dickinson, 1987; Dickinson, Nicholas, & Adams, 1983). For example, Dickinson (1980) reviews the sensory preconditioning paradigm (Nader & LeDoux, 1999; Rizley & Rescorla, 1972; Talk, Gandhi, & Matzel, 2002; Young, Ahier, Upton, Joseph, & Gray, 1998). During stage 1, a tone is paired with food. During stage 2, the food is paired with illness. During stage 3, appetitive response to just the tone on its own is tested. Rats trained on stages 1 and 2 show an attenuated response in stage 3 in comparison to rats that were trained on only stage 1. This demonstrates that rats are able to integrate the knowledge acquired in stages 1 and 2, and Dickinson interprets this (and other data) as evidence for the presence of declarative representations (i.e., an internal world model).

Our approach is a type of model-based reinforcement learning in which an agent learns to maximize reward through trial-and-error interaction with its environment. This satisfactorily captures the problem faced by a real animal. First, we assume that our model can recognize the current environmental stimulus and the time since its onset, by activating one of a set of predefined and fixed internal states. This assumption is similar to the tapped-delay-line representation assumption of Schultz et al. (1997), Montague et al. (1996), and others. Second, these states are built into an explicit internal model of the environment that comprises a transition function and a reward function. Third, the internal model is used to evaluate alternative actions and generate motivation via the online calculation of expected future reward. We will then propose a role for dopamine in modulating the efficacy of the connections between the states of the internal model, effectively implementing an online version of the discount factor of reinforcement learning (Sutton & Barto, 1998).

Figure 1 illustrates our tapped-delay-line assumption. The temporal resolution at which stimuli are represented is arbitrary, and we choose an interval of 1s. This interval will affect the quantitative but not the qualitative nature of the results. The availability of states is assumed to be sufficient for each task that is considered. Each state, s_i, is associated with an intrinsic reward value, $r(s_i)$, that is assumed to be supplied by the environment. For example, a state representing shock might be assumed to elicit a reward value of -1, while a state representing food might elicit a reward of 1. Neutral stimuli will always elicit a reward value of zero. We assume that a number of different actions are available to the agent, $A = \{a_1, a_2, ..., a_m\}$, where m is the total number of actions available. We also assume that the agent can take an action only when a new environmental stimulus is presented. For example, actions are not taken in $s_{\text{'Any'},t>0}$.

The internal model is described with reference to the simple neural circuit shown in Figure 2. Each time a new internal state, s_{new}, is activated, the

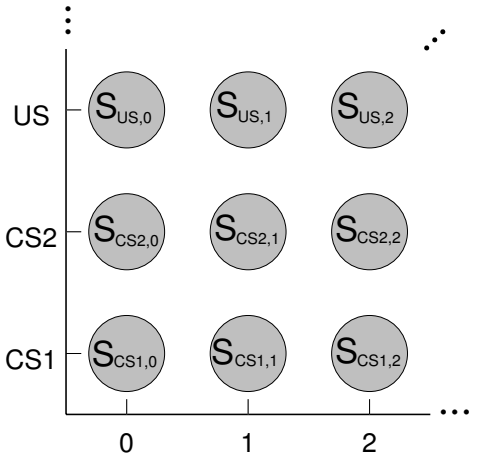


Figure 1: We assume that the model is able to recognize and represent the current stimulus and the time since its onset by activating one appropriate internal state unit from a predefined and fixed set, $S = \{s_1, \ldots, s_n\}$, where n is the total number of states. We will find it convenient to index each unit by either a stimulus-offset pair (as in the figure) or a single generic index that uniquely labels each state. For example, s_i is simply the i^{th} state.

following procedure is performed:

1. **Update reward estimate of the new state:**

   $$\hat{R}(s_{new}) := \hat{R}(s_{new}) + \alpha\big(r(s_{new}) - \hat{R}(s_{new})\big).$$

2. **Update the transition connections:**

   $$\hat{T}(s_{old}, a_{old}, y) := \begin{cases} (1-\beta)\hat{T}(s_{old}, a_{old}, y) + \beta & \text{If } y = s_{new}, \\ (1-\beta)\hat{T}(s_{old}, a_{old}, y) & \text{Otherwise.} \end{cases} \quad (3.1)$$

   for all states, $y \in S$.

3. **Action selection:** If s_{new} represents the onset of an external stimulus, then select and take action, a_{new}, as described below.

In the above, s_{old} is the previously active state and a_{old} is the previously selected action. α and β are learning rates, which are arbitrarily defined by $\alpha = \beta = \exp(-\frac{trial}{100})$, where *trial* is the trial number. These learning rates start at 1 and are slowly reduced to 0 across trials.

Some environmental states are marked as terminal states (see the environment descriptions later), and when a terminal state transition occurs, the

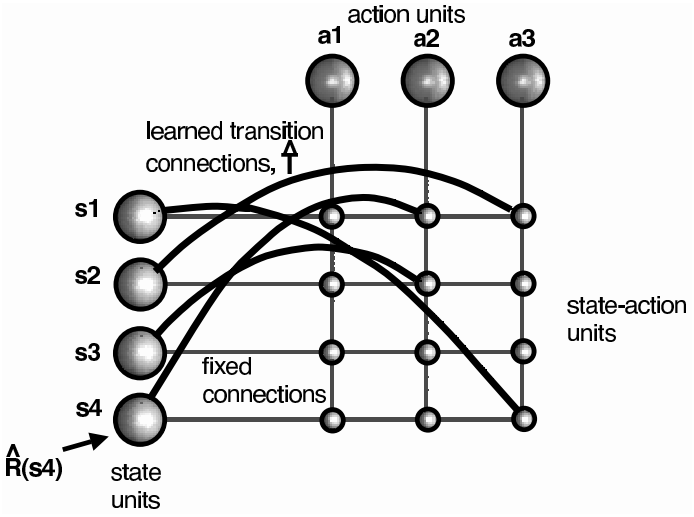


Figure 2: The internal model can be interpreted using a simple neural metaphor. Each internal state unit is connected to each other state unit under each possible action by a unique transition connection, the strength of which is controlled by a transition weight. For example, the transition strength, $\hat{T}(s_i, a_2, s_j)$, will be adapted during learning to reflect the probability that state s_j follows state s_i under action a_2. Note that the direction of the transition connections is from the state-action units back to the state units. Additionally, each state maintains an estimation of the immediate reward value associated with that state, $\hat{R}(s \in S)$. Although this value will be immediately available in $r(s \in S)$, the former represents the estimate learned by the internal model, while the latter represents the actual value coming in from the environment. In theory, $r(s \in S)$ could be a distribution, for example, while $\hat{R}_(s \in S)$ is always a scalar value. The reward values and transition connections are adapted during learning so that an explicit internal model of the environment is approximated. This model will be used to estimate future reward and to drive motivation and action selection. From an anatomical perspective, we speculate that the states themselves are represented in cortical areas (possibly including the OFC and amygdala), while the transition connections are routed through the ventral striatum (see the discussion).

trial is ended. Over a number of trials, the agent learns to represent an explicit internal model of its environment that consists of an estimated reward function and an estimated transition function. Step 2 just redistributes the weights from the relevant state-action unit (see Figure 2) back to each state unit according to a learning rate, β. All transition strengths are initialized to $1/n$, and the redistribution rule ensures that $\sum_{i=1}^{n} \hat{T}(x, y, s_i) = 1$ for all $x \in S$ and $y \in A$ at all times.

Action selection is performed by using a look-ahead process to generate the expected future reward of each action in turn and then selecting the best action. The look-ahead process involves playing through the consequences of taking each action inside the internal model. We assume that this look-ahead can be performed without disturbing the modeling process described above. If actually implemented in neural circuitry, it might be necessary to maintain two copies of the internal model: one for keeping track of the environment and one for performing look-ahead. We abstract over this detail. We now propose a role for dopamine in modulating the efficacy of the transition connections. First, let z_{ij} denote the state-action unit corresponding to taking action j in state i. Next, let the activation of s_i and z_{ij} be denoted by $\xi(s_i)$ and $\xi(z_{ij})$ respectively. Also, let $FutRew(a \in A)$ be an internal register used for accumulating the total future reward of taking action a in current state, s_{new}. The action selection step is fleshed out as follows:

3. Action Selection: If s_{new} represents the onset of an external stimulus:

   (a) For each action, $a_i \in A$:

      i. $FutRew(a_i) := 0$

      ii. For all $j \in \{1...n\}$:
      $$\xi(s_j) := \begin{cases} 1 & \text{If } s_j = s_{new}, \\ 0 & \text{Otherwise.} \end{cases}$$

      iii. For some fixed number of iterations, q:

         A. Propagate activation from state units to state-action units:
         For all $j \in \{1...n\}$ and all $k \in \{1...m\}$:
         $$\xi(z_{jk}) := \begin{cases} \xi(s_j) & \text{If } k = i, \\ 0 & \text{Otherwise.} \end{cases}$$

         B. Generate hypothetical next state:
         For all $j \in \{1...n\}$:
         $\xi(s_j) := \sum_{k=1}^{n} \sum_{l=1}^{m} \xi(z_{kl}) \times \hat{T}(s_k, a_l, s_j) \times DA_{tonic}$

         C. Collect rewards for this hypothetical state:
         $FutRew(a_i) := FutRew(a_i) + \sum_{j=1}^{n} \xi(s_j) \times \hat{R}(s_j)$

         D. Return to 3(a)iiiA.

   (b) Select and take action,

$$a_{new} := \begin{cases} \underset{a \in A}{\text{argmax}}\, FutRew(a) & \text{With prob. } p, \\ \text{Random action} & \text{With prob. } 1-p. \end{cases}$$

In the above, $0 \leq DA_{tonic} \leq 1$ is the level of tonic dopamine (default = 1); q is the maximum depth of the look-ahead process, which we arbitrarily set at 50; and p controls exploration, where $p = (1 + \exp(5 - 0.07trial))^{-1}$. Exploration starts at 1 and is reduced to 0 over successive trials. Exploration is an important part of behavior acquisition, since in general the initial state of the internal model will not accurately reflect the environmental contingencies.

The approach we have taken is very simple. Activity cycles around the circuit of Figure 2 in order to simulate the environmental consequences of each action in sequence. The action that accumulates the greatest future reward during this process is actually selected, generating a new real sequence of environmental stimuli. A role for DA_{tonic} as a modulator of the transition connections is proposed in step 3(a)iiiB. This will allow us to account for the finding that actions motivated by distal rewards are more vulnerable to dopamine manipulation than those motivated by proximal rewards. We have made the simplifying assumption that during the look-ahead process, the action currently being evaluated is always selected (step 3(a)iiiA). We consider a more flexible alternative to this look-ahead policy later, although the qualitative nature of the results are not contingent on this feature of the model because of the simple environments used.

Providing that $DA_{tonic} = 1$, then at any stage in the look-ahead process, $\sum_{j=1}^{n} \xi(s_j) = 1$. This is guaranteed by the starting condition in 3(a)ii and the fact that $\sum_{j=1}^{n} \hat{T}(x, y, s_j) = 1$ for all $x \in S$ and $y \in A$. More important, the activation of each state unit corresponds to the probability of that state indeed being the current state at that future time, under the look-ahead policy. However, for $DA_{tonic} < 1$, these activations will decay during the look-ahead process, although the activations will still be in proportion to these probabilities. As the look-ahead process continues, $FutRew(a)$ converges on the estimated future reward of taking action a, modulated by the online discount factor, DA_{tonic}. Expected future reward or the return is the standard quantity to maximize in reinforcement learning problems (Sutton & Barto, 1998; Kaelbling, Littman, & Moore, 1996). Note that the agent is naturally parallel in design and relies mainly on local learning and activation rules.

4 Modeling Conditioned Avoidance

A prerequisite of any model of delusions or disordered thoughts is a model of ordered thoughts. Since a model of ordered thoughts represents the holy grail of a number of disciplines, it is as well that we have a relatively simple yet highly relevant animal model of dopaminergic action at our disposal.

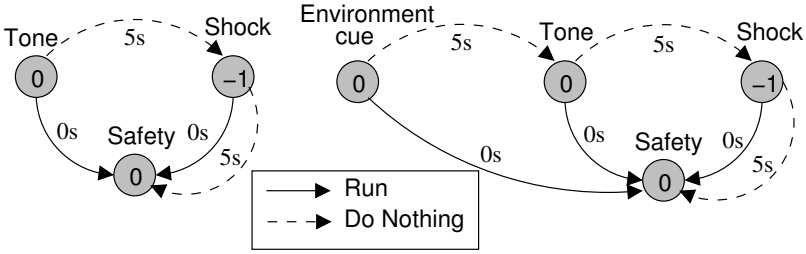
We present the model in conjunction with a generalized version of the CA paradigm presented in Maffii (1959) (reviewed with other classic APD studies in Dews & Morse, 1961). The standard CA finding that APDs disrupt avoidance before escape is observed within Maffii's paradigm, but Maffii

also makes an interesting additional observation that we wish to model. He found that after sufficient training, the rats began producing the avoidance response as soon as they were placed in the cage and before they were even presented with the tone. He termed this response to environmental context the *secondary avoidance response*, and the standard response to the tone the *primary avoidance response*. He then found that not only was the primary avoidance response more vulnerable to dopamine blockade than the escape response itself (the standard finding), but that the secondary avoidance response was more vulnerable than the primary response (see Figure 5, left). Apparently, the more distal the cue from the shock, the more vulnerable that cue was to dopamine blockade in terms of its ability to elicit a response.

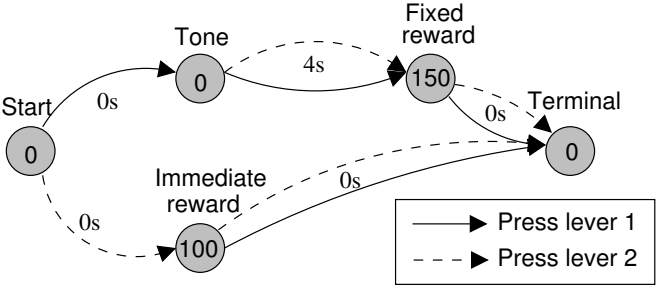
We can describe the standard CA environment with the finite state model of Figure 3a (left) and Maffii's experiment with Figure 3a (right). The finite state environments defined in Figure 3 are a necessary assumption required to formalize the problem, and they effectively take on the role of an animal's environment as well as its sensory processing. These environment descriptions will be different for each experiment that we consider. In contrast, the learning rules described above are defined once and represent our model of the behavioral processes, which we believe to pertain to the ventral striatal D2 receptor.

Figure 4(a) shows the internal model after 100 learning trials. The transition connections reflect the environmental contingencies of Figure 3a (right). This internal model can then be used to generate the expected future reward of taking different actions in each environment state. For example, Figure 4e shows how activation is propagated through the internal model during the look-ahead process of the "do nothing" action at trial onset for $DA_{tonic} = 1$ for the internal model of Figure 4c. The return is generated by summing the reward estimates of each active unit in proportion to its activation. Setting $DA_{tonic} < 1$ after acquisition of the internal model results in the decay of this activation during the look-ahead process, and therefore also the attenuation of the impact of increasingly distal outcomes (not shown).

Figure 5 compares expected future reward thus generated with the performance of Maffii's rats. Here we are assuming that expected future reward is a suitable basis for motivation. In order to convert this value into the probabilistic behavior of the rats as shown in Figure 5 (right), a softmax action selection mechanism would be required in place of the ϵ-greedy mechanism of step 3b. Although not shown, the model acquires avoidance behavior within a number of trials that is consistent with the amount of experience required by an actual rat (Wadenberg et al., 2001). However, this behavior is rather trivial, and the real point of interest is the selective effect of DA_{tonic} on secondary avoidance versus primary avoidance versus escape. For interest, Figures 4c and 4e give an example of the model's behavior in noisy or stochastic environments.



(a) MDP for modeling Conditioned Avoidance



(b) MDP for discounting task of (Wade et al., 2000)

Figure 3: (a, left) A simple and abstract environment model for the classic CA task. Each circle represents a different environment state an agent can be in, and the arrows denote the outcome of each of the two possible actions. The arrows are labeled with the time between the action being taken and the transition actually occurring. For example, if the Do Nothing action is selected when the tone is presented, the shock will be delivered after a delay of 5 s (see Wadenberg et al., 2001, for an experimental example). We make the simplifying assumption that the agent must take one of the available actions on entry into a state, and must then wait for the ensuing transition before another action may be taken. The number inside each state represents the reward, r, associated with that state. The Safety state is the terminal state, which always transitions to itself, and the trial is terminated when the first such transition is made. Note that if the agent selects Do Nothing in the Shock state, the shock lasts only for 5 s, after which the trial is automatically terminated. This is consistent with a typical experimental setup (Wadenberg et al., 2001). (a, right) A generalized version of (left) that represents the experimental setup of Maffii (1959). (b) An abstract definition of the environment for the discounting task of Wade et al., (2000; see section 5). Trials begin in the left-most state.

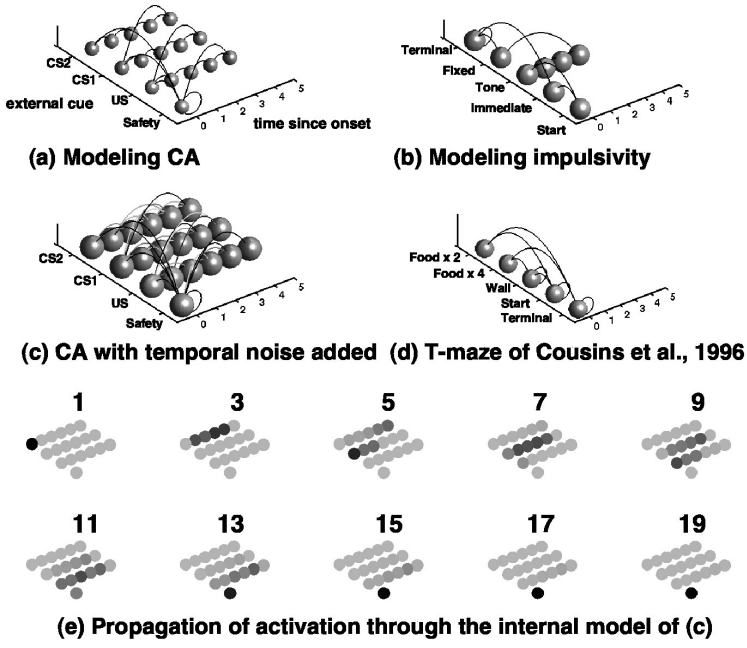


Figure 4: (a) The state of the internal model after learning the environment of Figure 3a (right). The secondary stimulus refers to the Environment cue in Figure 3a (right) and the primary stimulus to the Tone. All learned transition connections, $\hat{T}$, are shown. The relationship to Figure 2 is as follows: The units of Figure 2 are plotted here in weight space rather than physical or anatomical space. Also, the state-action units are removed, and the transition connections from each of the state-action units are collapsed onto the relevant state so that the connections from state to state can be seen more clearly. As a result of visualizing the $\hat{T}(s, a, s')$ function as a simpler $\hat{T}(s, s')$ function, it is not possible in this figure to know which action causes a given transition. However, the transition connections associated with taking the Run action in response to the onset of each external stimulus always transition to the Safety unit. The remaining transitions show the consequences of Do Nothing. (b) The state of the internal model after learning the environment of Figure 3b. (c) As in *a* except that temporal random noise is added to the identification of the current state. The bolder connections denote a stronger transition weight. (d) The state of the internal model after learning the T-maze environment of Figure 8(a). (e) The propagation of activation through the internal model of *c* during a typical look-ahead process for the Do Nothing action at the beginning of a trial, with $DA_{tonic} = 1$. The process is shown for every alternate iteration.

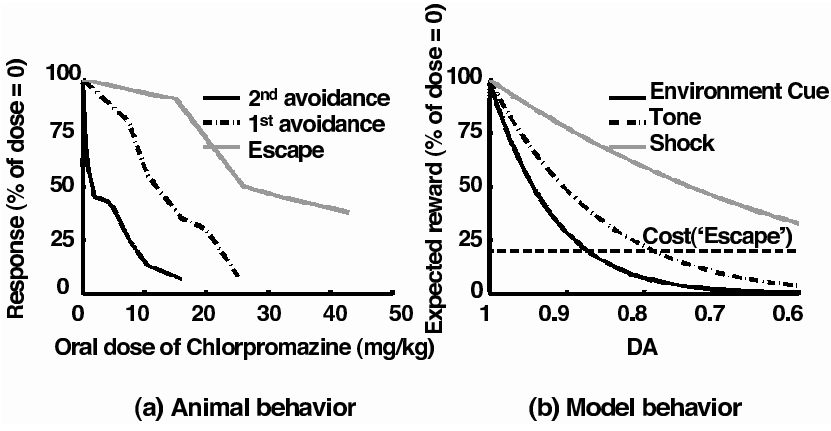


Figure 5: A comparison of model performance with Maffii's data. (a) Number of secondary avoidance responses, primary avoidance responses, and escape responses under increasing doses of the neuroleptic chlorpromazine (a drug with a particularly high affinity for the D2 receptor), as a percentage of the number of responses without the drug (adapted from Maffii, 1959). (b) The change in *FutRew*(Do Nothing) for each of the three important states (solid line = environment cue, dashed line = tone, gray line = shock) as DA_{tonic} is decreased. Since we use DA_{tonic} as an abstract representation of dopamine and the model does not attempt to address the underlying neurochemical processes, the relationship between the model parameter, DA_{tonic}, and chlorpromazine dose is uncertain. However, it is the qualitative nature of the results that is of interest, and in particular, the selective effect of dopamine blockade on secondary avoidance versus primary avoidance versus escape. For the comparison to be meaningful, we assume that *FutRew*(Do Nothing) can be used as a direct analogy of motivation, since the alternative action, Run, always yields an estimated future reward of zero. The horizontal line suggests an example escape cost that could be used to threshold motivation. This could explain why many studies find that low doses of neuroleptics disrupt avoidance but not escape.

5 Impulsivity, Delayed Rewards, and ADHD

Attention deficit/hyperactivity disorder (ADHD) is a developmental disorder affecting 3% to 7% of school-age children, characterized by inattention, hyperactivity, and impulsivity (Frances, 2000). The single most effective treatment for the disorder is medication with psychostimulants such as methylphenidate (Phares, 2003). Psychostimulants (of which amphetamine is one) are known to increase extracellular concentrations of dopamine (Seeman & Madras, 1998), and indeed ADHD is believed to be primarily a dopaminergic/noradrenergic disorder (see Phares, 2003). Decreased blood flow has also been observed in the striatum of ADHD patients, a deficit

reversible by treatment with methylphenidate (reviewed in Schneider, Sun, & Roeltgen, 1994). As with psychosis, striatal dopamine seems to be of importance to ADHD. However, unlike psychosis, it may be a deficit rather than an excess that is to blame.

There are no laboratory tests, neurological assessments, or attentional assessments that have been established as a diagnostic in the clinical assessment of ADHD (Frances, 2000). However, Solanto et al., (2001) have suggested that a desire to avoid delay (the delay aversion hypothesis) is one of the best characterizations of impulsivity with respect to ADHD, and Catania (in press) has argued specifically that many of the symptoms of ADHD can be accounted for by assuming too steep a discounting gradient. Studies such as de Wit et al. (2002) have confirmed that certain measures of impulsivity are indeed reduced in healthy volunteers by amphetamine. For example, subjects were asked questions such as, "Would you prefer ten dollars in thirty days or two dollars at the end of the session?" The amphetamine-treated subjects showed an increased preference for the larger but delayed reward when compared with a placebo group.

If blocking dopamine can disrupt avoidance responding in rats, enhancing dopamine seems to decrease impulsivity as measured by this kind of delay discounting (Wade et al., 2000; Richards et al., 1999; de Wit et al., 2002). Furthermore, both phenomena may be mediated by the same part of the brain: the striatum (see Phares, 2003, for more on striatal involvement in ADHD). However, as with psychosis, the underlying behavioral and psychological processes are unclear. In a bid to better understand the role of dopamine in impulsivity and ADHD, Richards et al. (1999) and Wade et al. (2000) have investigated the effects of both amphetamine- and dopamine-blocking drugs on delay discounting in rats. Their experimental paradigm can be summarized as follows.

A thirsty rat is trained to press one of two levers. One lever yields an immediate reward (for example, 100 μl of water), and the other yields a fixed reward (150 μl of water) but only after a delay of 4 s. If the animal selects the delayed reward, a tone is presented between the lever press and the water to make the task easier. The immediate reward is then adjusted from trial to trial depending on which lever the animal selected on the previous trial. If the rat chose the immediate reward, the immediate reward is reduced by 15%, and if the rat chose the delayed reward, the immediate reward is increased by 15%. In this way, the immediate reward is varied until the rat has no particular preference. The amount of immediate reward elicited at this indifference point is then interpreted as the animal's "value" of the fixed, delayed reward. The rats are trained over a period of many weeks, on a variety of different starting conditions, until they become familiar with and adept at exploring the two alternatives and achieving the indifference point that suits their preference. As an additional aid, if the rat chooses the same lever twice in a row, then the immediately following trial is a forced exploration trial in which only the other lever yields a reward.

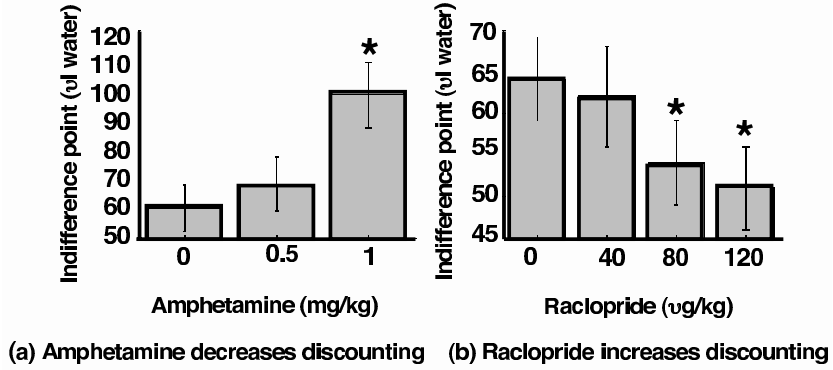


Figure 6: The impact of various doses of (a) a dopamine-enhancing drug (amphetamine) and (b) a D2-blocking drug (raclopride), on the indifference point. The indifference point indicates the value of the delayed 150 μL alternative. Error bars indicate SEM, and asterisks denote a significant difference from dose = 0. Adapted from Wade et al. (2000).

After the rats have been trained on this procedure, they are tested under various systemic doses of both amphetamine and raclopride (dopamine D2 receptor blocker; see Figure 6). A dose-dependent effect of both drugs is observed on the indifference point. They claim that amphetamine has effectively reduced impulsivity (by increasing the value of the delayed reward), and raclopride has increased impulsivity (by decreasing the value of the delayed reward). Because of the incrementally adjusting procedure used, it is impossible to rule out a learning effect completely. However, the effects of these drugs were very quick compared with the weeks of training required for acquisition of the basic task. In the case of raclopride in particular, the lower indifference point was reached as quickly as the paradigm allowed—after just a few trials under the drug. It therefore seems likely that a significant part of the change in the animals' behavior was due to the effect of drug on performance (rather than or in addition to its effect on learning).

Wade et al. (2000) conclude that "this pattern of results indicates that blocking D2-receptors may have a selective effect on the value of delayed rewards" (p. 197). We can now use the model described in the previous section to propose an explanation that is consistent with data from CA. Figure 4b shows the result of training the agent on the environment of Figure 3b. Acquisition is again trivial (for $DA_{tonic} = 1$, the agent learns to select the greater, delayed reward), but Figure 7 shows the effect of manipulating dopamine after acquisition. The delayed reward is discounted more in

the look-ahead process as DA_{tonic} is reduced. In order to capture the effects of both dopamine blockade and amphetamine, we assume a baseline $DA_{tonic} < 1$. The model is then able to provide a qualitative account of all drug-induced changes in impulsivity. Note that the vertical bars shown in Figure 7 can be moved apart (or together) by reducing (or increasing) the temporal resolution in Figure 1, thereby modifying the steepness of discounting. For reference, the largest dose of raclopride used (120 μg) would reduce avoidance responding in CAR by around 20% (Wadenberg, Kapur, Soliman, Jones, & Vaccarino, 2000).

Notionally, the model also captures five additional observations made by Wade et al. (2000). First, independent of drug treatments, they vary the delay to the fixed reward (2 s and 8 s). They find that a delay of 2 s increases the indifference point (value of the delayed alternative) and that a delay of 8 s decreases the indifference point. Second, they find that the initial amount of water on the immediate alternative does not affect the eventual indifference point. Third, they find that the deprivational state of the animal (i.e., its thirst) does not affect the indifference point. In our model, a logical role for thirst would affect the perception of the two rewards equally (i.e., by a common factor), which could be achieved, for example, by substituting step 3(a)iiiC with:

3(a)iiiC) *Collect rewards for this hypothetical state.*
$FutRew(a_i) := FutRew(a_i) + \sum_{j=1}^{n} \xi(s_j) \times \hat{R}(s_j) \times$ "Thirst".

This would not affect the relative indifference point. Fourth, our model predicts that blocking dopamine should generally reduce the motivation to press either lever, and enhancing dopamine should generally increase the motivation to press a lever. Wade et al (2000) examined the tendencies of the rats to complete each trial and found that raclopride reduced but amphetamine increased the mean number of trials completed. Admittedly, completed trials is only an informal measure of motivation. Also, in a related and recent study, J.B. Richards (personal communication, 2003) finds that if rats are trained on a task in which levers yield immediate rewards, but one is certain and one is uncertain, then amphetamine does not affect the indifference point. This behavior would be produced by the model too because uncertainty is represented by the transition connection strengths, and dopamine has an equal effect on all such connections. A complete analysis of the model's performance under these five conditions is outside the scope of the account presented here.

It should be noted that Wade et al. (2000) observed the effects described above only with dopamine D2-blocking drugs. D1 receptor blockers had no significant effect. Interestingly, Cardinal et al. (2001) demonstrate that accumbens (core) lesions in rats cause exactly the same kinds of effects, leading them to conclude that the accumbens is involved in the pathogenesis of impulsive choice, a finding they suggest can shed light on ADHD, addiction, and other impulsive control disorders. (Cardinal, Robbins, & Everitt (2000)

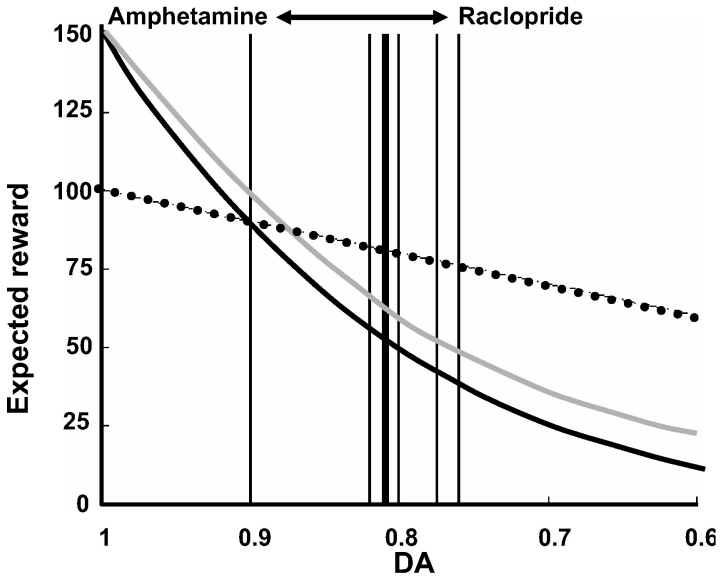


Figure 7: The effect of varying the DA_{tonic} parameter after learning on the estimated future reward associated with each of the two actions from the Start state. The solid curve shows the estimated future reward of pressing lever 1 (delayed reward =140 μL), while the horizontal dashed line shows the estimated future reward of pressing lever 2 (immediate reward = 100 μL). The gray curved line shows the hypothetical value of the immediate reward necessary to balance the estimated future reward of the immediate alternative with that of the delayed alternative (i.e., the indifference point). This gray line can be used to match the model performance with data from Wade et al. (2000). The thick vertical bar shows the simulated level of dopamine that corresponds to the performance of the undrugged rats in Wade et al. (2000). The vertical bars to the left show the simulated dopamine level corresponding to the indifference point of the amphetamine-treated rats, and the vertical bars to the right show the simulated dopamine level corresponding to the performance of the raclopride-treated rats (see Figure 6). In agreement with the animal study, the model predicts that reducing DA_{tonic} reduces the indifference point, while increasing dopamine increases the indifference point.

also find the same kinds of impulsive behaviour with systemic administration of amphetamine and combined D1/D2 blockers.

It has been suggested that accumbens dopamine is necessary for producing anticipatory responses (e.g., avoidance, or lever pressing for food or water), but not consummatory responses (e.g., escape, or feeding or drinking itself) (Ikemoto & Panksepp, 1999). It is therefore noteworthy that in the impulsivity studies discussed above, two equally anticipatory responses with

equal motoric requirements were dissociated by dopamine manipulation. Therefore, arguments that rely solely either on anticipatory-consummatory distinctions or on motor deficits may need to be adjusted to address delay-discounting paradigms.

Finally, two important caveats are considered. First, in contrast to the data discussed above, some studies have observed that amphetamine actually increases impulsivity rather than decreases it, leading to the suggestion that the finer experimental details, such as whether a cue is presented during the delay, may be important (Cardinal et al., 2000). Perhaps one problem is that amphetamine injections directly into the accumbens increase general locomotor activity (reviewed in Pennartz, Groenewegen, & Silva, 1994), as well as an animal's motivation to work for reinforcers (Taylor & Robbins, 1984), as predicted by the internal model since expected future reward equals motivation). Therefore, it may be difficult for impulsivity studies to separate the absolute increases in motivation due to amphetamine from the relative decreases in choice tasks. The way in which the animal internally models its environment is likely to play a large role. Second, it must be acknowledged that ADHD is a complex and multidimensional disorder, and the delay aversion hypothesis (see Solanto et al., 2001) and delay discounting hypothesis (see Catania, in press), represent only one line of argument. However, we conclude by contrasting our proposed role for dopamine in gating a look-ahead process with one of the standard characterizations of impulsivity within ADHD offered by DSM-IV (emphasis our own): "Impulsivity may lead to accidents and to engagement in potentially dangerous activities *without consideration of possible consequences*" (Frances, 2000, p. 86).

6 A T-Maze Experiment

In a fascinating study by (Cousins et al. (1996), rats were presented with a choice between two arms of a T-maze: one leading to four food pellets, and the other to two. However, the arm leading to four pellets was obstructed with a barrier that had to be climbed (see Figure 8b). Once trained, the rats chose the arm containing four pellets on almost 100% of trials. However, after bilateral intra-accumbens injections of 6-hydroxydopamine, a treatment that destroys dopaminergic projections to the accumbens, the rats changed their behavior, selecting the unobstructed but lesser reward instead (80% of the time). A second experiment was then performed in which different rats were trained (untreated, as before) on a different version of the T-maze in which there was no reward in the unobstructed arm. Subsequent dopamine lesions only slightly reduced responding for the obstructed arm (from 100% of the time to 80%). Therefore, Cousins et al. (1996) were able to rule out the possibility that the animals in the first experiment were unable to cross the barrier because of motoric deficits for example. Rather, it appears that dopamine lesions were influencing the motivational and choice processes of the animal.

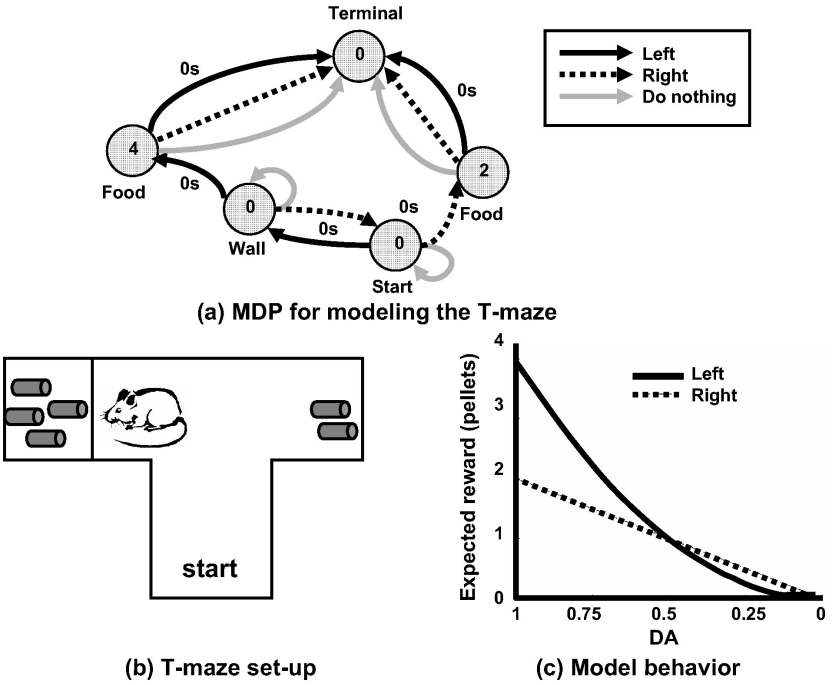


Figure 8: (a) A formal environment model capturing the salient features of the T-maze task. For simplicity we can assume negligible delays between all state transitions. This is not a necessary assumption. (b) The T-maze experimental setup used in Cousins et al. (1996). (c) The effect of varying DA_{tonic} ,after acquisition, on the estimated future reward associated with taking the Left (solid) and Right (dotted) actions from the Start state. The point at which the graphs cross denotes the point at which the model will switch from seeking the obstructed reward to seeking the easily available alternative.

The internal model can be used to capture these results by training it on the environment of Figure 8a. Figure 4d illustrates the resulting internal model, and Figure 8c demonstrates the effect of modulating DA_{tonic}, after acquisition, on the estimated future reward associated with moving left or right from the start state. The model accounts for the change in animal behavior from the obstructed to the unobstructed arm under dopamine disruption. It also notionally accounts for the observation that if the rats are trained with no food in the unobstructed arm, then they continue to select

the obstructed arm, although apparently with less motivation. The model would continue to select the obstructed arm in this case until the solid line failed to exceed the cost of climbing the wall.

The results of this study are particularly striking because the nucleus accumbens is specifically targeted with a direct dopamine challenge, eliciting a qualitative change in behavior. This result is not an isolated experiment either. It is duplicated in essence in Salamone et al. (1991, 1994) with both direct accumbens dopamine depletion and systemic D2 blockers, and Salamone et al. (1997) provides a review of many other related experiments by Salamone and colleagues that adhere to the same general pattern.

7 Instrumental Responding for Rewards

A generally accepted consequence of dopamine blockade in rats is a reduction in lever pressing (also other responses) for various types of reward (Ettenberg, Koob, & Bloom, 1981; Evenden & Robbins, 1983; Fibiger et al., 1976; Fowler et al., 1986; Rolls et al., 1974; Salamone et al., 1993; Wise & Schwartz, 1981; Wise et al., 1978). Moreover, lever pressing for food or water is often found to be reduced at doses that have little or no impact on free feeding or free drinking where no lever press is required (Rolls et al., 1974; Salamone et al., 1993; see Figure 9a). However, recall that Berridge and Robinson (1998) were able to disrupt free feeding, even if the food was under the rats' noses, by reducing accumbens and neostriatal dopamine by around 95%. We can use the current model to account for these findings by assuming the environment model of Figure 9b. In order to estimate future reward (and therefore generate motivation) for lever pressing, all three of the transitions are required. In contrast, free feeding requires only Approach and Consume, and in Berridge's rats' case, only the Consume transition is required. We do not show the model's performance on this task because of the qualitative similarity between Figures 9a and 5a. However, it should be evident that lever pressing is disrupted before free feeding and free feeding before consumption in the model under the assumption of Figure 9b.

With respect to this and the T-maze experiment reviewed earlier, it is an open question as to whether it is the "instrumental" or temporal distance (or both) between an action and its rewarding outcome that renders that action vulnerable to dopamine blockade.

8 Uniting Existing Hypotheses and Related Work

We have presented a computational model intended to address a corpus of experimental data that point to a selective role for dopamine in the expression of previously acquired behaviors. More specifically, the D2-receptor subtype within the ventral striatum (particularly the accumbens) has emerged as a common neuroanatomical thread throughout the studies we have considered. If our model of the transition connections passing

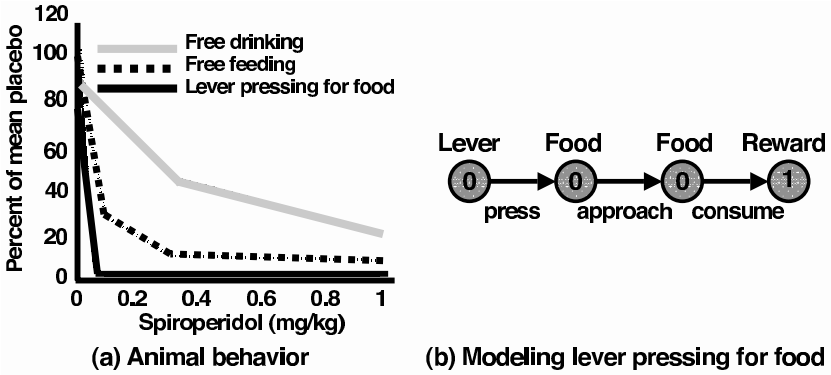


Figure 9: (a) Lever pressing for food is attenuated at neuroleptic doses that do not drastically affect free feeding. Free drinking may be even more robust than free feeding to dopamine challenge (results adapted from Rolls et al., 1974). Note that spiroperidol primarily blocks the D2 receptor. (b) An abstract environment for instrumental responding that, in combination with the model presented, accounts for the selective vulnerability of instrumental responding to dopamine challenge.

through the ventral striatum is plausible, then we would expect to see this area being activated in response to conditioned stimuli as the look-ahead process is invoked. A number of fMRI studies in human subjects confirm this for both appetitive (Gottfried, O'Doherty, & Dolan, 2002; O'Doherty, Deichmann, Critchley, & Dolan, 2002; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003) and aversive (Jenson et al., 2003) tasks. Also, using electrophysiological recording techniques, Schultz, Tremblay, & Hollerman (2000) have found a variety of activations in the ventral striatum that are consistent with its role in the expectation of reward across a delay.

In addition to the ventral striatum, a range of fMRI, electrophysiological, and behavioral data points to a central role for the amygdala (Cador, Robbins, & Everitt, 1989; Everitt, Morris, O'Brien, & Robbins, 1991; Everitt & Robbins, 1989; Cardinal et al., 2002; Nishijo, Ono, & Nishijo, 1988; Robbins, Cador, Taylor, & Everitt, 1989; Tremblay & Schultz, 2000a, 2000b) and the orbito-frontal cortex (Arana et al., 2003; Bechara, Damasio, Tranel, & Anderson, 1998; Cardinal et al., 2002; Iversen & Mishkin, 1970; Masterman & Cummings, 1997; O'Doherty et al., 2003) in the representation of environmental reward contingencies of the kind relevant to the current discussion. It has been suggested that the OFC in particular is crucially involved in the motivational control of goal-directed behavior (Schultz et al., 2000) learning about rewards (Dias, Robbins, & Roberts, 1996), and evaluating alternatives

(Arana et al., 2003; Schultz et al., 2003) via a common neural currency (Montague & Berns, 2002). We therefore speculate that the states themselves may be represented cortically (originating presumably in the sensory cortex), with the reward estimates, $\hat{R}$, along with the the return being represented in the OFC and amygdala.

The proposed model can be used to unite a number of existing cognitive hypotheses of dopamine function within a formal framework. For example, Berridge and Robinson (1998) suggest that mesolimbic dopamine mediates the wanting component of reward as distinct from the liking, and in our model, following McClure et al. (2003), we interpret expected future reward as precisely this wanting. Salamone et al. (1997) suggest that "accumbens dopamine is important for responding to stimuli that are spatially and temporally distant from the organism" (p. 353). This statement precisely summarizes the psychological value of dopamine in our model. Ikemoto and Panksepp (1999) have also noted the relevance of the proximal-distal distinction to mesolimbic dopamine function. They argue that nucleus accumbens dopamine is important for invigorating flexible approach responses (distal), as distinct from consummatory responses (proximal). With respect to our model, and particularly section 7, it should be noted that consummatory behaviors such as free feeding may not be disrupted with only accumbens dopamine depletions. For example, Berridge and Robinson (1998) achieved aphagia with both accumbens and neostriatal dopamine reduction. The striatum may therefore be functionally differentiated, with the ventral region specializing in distal motivation (e.g., approach) and the dorsal region in proximal motivation (e.g., consumption). Modeling this distinction must constitute future work.

Our approach is different from the TD prediction-error hypothesis (see Houk et al., 1995; Schultz et al., 1997; Montague et al., 1996), which suggests that the phasic dopamine response signals the difference (error) between the future reward predicted by the animal and the actual reward received. This error is then used to drive the learning process in a biologically plausible fashion (Waelti, Dickinson, & Schultz, 2001). In contrast, we have proposed a role for tonic dopamine in the generation of expected future reward that is independent of the acquisition process. The advantages of our approach are that we can model the effect of dopamine manipulation not only on the expression of previously acquired behaviors, but also the sensitivity of this effect to the relationship between action (or CS), and outcome (or US).

A weakness of our internal model is that it fails to address the role of dopamine in the acquisition process or the phasic response of dopamine neurons themselves. We therefore suggest that future work should be oriented around hybrid approaches aimed at achieving a more comprehensive account of the neuromodulator. Toward this end, a number of discussions and models have been proffered that extend TD-based representations with

explicit internal model representations (Daw, Courville, & Touretzky, 2004; Dayan, 2002; Dayan & Balleine, 2002; Suri & Schultz, 1998; Suri et al., 2001; Suri, 2001, 2002). However, a significant challenge remains in bridging the gap between models of dopamine neuron firing and models of behavioral and psychological phenomena in which dopamine may play a pivotal role. It is particularly important that we achieve a better understanding of whether explicit internal model representations or a cached value function (as in TD) is most appropriate for modeling the brain reward system.

To conclude this section, we speculate as to why the brain might need DA_{tonic}. Some suggestions include (1) constraint of the look-ahead process via its action as an online discount factor; (2) adaptation of the trade-off between proximal and distal rewards in response to environment cues (Wilson & Daly, 2004) or deprivational states (Giordano et al., 2002); or (3) global control of general motivated activity.

9 Model Predictions and Future Work

Model is only as good as the useful predictions it makes, and so we are actively engaged in a program of experimental validation. In the account presented above, we allowed the model to make an action choice only in response to the onset of an external stimulus. However, within the context of conditioned avoidance, we have also looked at the predictions made by the model if an action can be selected in any state, including the internal timing states. These novel model-driven predictions were subsequently validated with experimental data (Smith et al., 2004), leading us to argue against the preeminent motor deficit hypothesis (Aguilar et al., 2000; Ogren & Archer, 1994) in favor of a motivational hypothesis of APD-induced avoidance disruption in rats. We are looking for interactions of dopamine manipulation with CS-US interval within CA, with a view to further testing the internal model account of motivated behavior.

One of the primary motivations for undertaking this work is to create a computational dopamine hypothesis that can be used to shed light on schizophrenia. We argue that many of the negative symptoms of schizophrenia can be notionally captured by reducing DA_{tonic}, including reduced motivation, flat affect, and anhedonia (given the suggestion that many of the rewards of life may actually be conditioned stimuli; Wise, (2002). Grace (1991) and Moore, West, and Grace (1999) have argued that a tonically hypo-dopaminergic state may be a key step to the development of psychosis, with the latter perceived as a hypersensitivity to phasic dopamine caused as result of homeostasis to the former. With respect to modeling psychosis, we suggest that an aberrant dopamine signal could lead to the construction or modulation of an aberrant internal model, and an aberrant internal model seems to be an excellent starting place to model delusions. Further research into combining a phasic (learning) role for dopamine with the internal model would be expected to flesh out this hypothesis.

One model assumption deserves a brief revisit. We enforced a very simple policy in the look-ahead process (step 3(a)iiiA), in which the current action being evaluated was always selected at each subsequent hypothetical state. This approach was simple and adequate given the way in which the problems were formulated. However, a more general and flexible look-ahead process would search different possible actions as one might in a game of chess, for example. However, since a branching search process is potentially costly, an alternative is to represent a Q-value at each state-action unit and then use this value to guide the search process during look-ahead. Such a value could be constructed using either a TD or Monte Carlo method (see Sutton & Barton, 1998). This would allow the agent to trade off the benefits of being able to explicitly simulate the consequences of actions (using the internal model) against the efficiency of simply using a precalculated value (an action-value function). The former is particularly important for being able to modulate behavior *after acquisition* based on an internal drive (e.g., salt deprivation) that was not present during conditioning. Berridge and Schulkin (1989) have demonstrated just such an ability in animals. There are already open lines of investigation into when and where internal model versus TD-like value functions influence motivated behavior (Dayan, 2002; Dayan & Balleine, 2002). However, using an action-value function to guide the online look-ahead process does not have a direct impact on the current discussion of dopamine function, which is kept as simple as possible.

In conclusion, we have demonstrated that an internal model approach is able to account for a range of experimental evidence that suggests that ventral striatal dopamine D2-receptor manipulation selectively modulates motivated behavior for distal versus proximal outcomes. Whether an internal model or the cached values of the TD algorithm are better placed to model both a tonic and phasic dopamine response in this brain region is likely to have important implications for understanding a number of human disorders, including schizophrenia and ADHD.

Acknowledgments

This work was primarily supported by an OMHF Special Initiative grant and a NET grant from the Canadian Institutes of Health Research. S.K. is additionally supported by a Canada Research chair. Thanks also to Ming Li and Jimmy Jensen for assisting in the review of the conditioned avoidance and fMRI literature (respectively) and to Christopher Fiorillo for reviewing an early draft of this work.

References

Ader, R., & Clink, D. W. (1957). Effects of chlorpromazine on the acquisition extinction of an avoidance response in the rat. *J. Pharmacol. Exp. Ther., 131*, 144–148.

Aguilar, M. A., Mari-Sanmillan, M. I., Morant-Deusa, J. J., & Minarro, J. (2000). Different inhibition of conditioned avoidance response by clozapine and Da D1 and D2 antagonists in male mice. *Behav. Neurosci., 114*(2), 389–400.

Anisman, H., Irwin, J., Zacharko, R. M., & Tombaugh, T. N. (1982). Effects of dopamine receptor blockade on avoidance performance: Assessment of effects on cue-shock and response-outcome associations. *Behavioral and Neural Biology, 36*, 280–290.

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., & Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *Journal of Neuroscience, 23*(29), 9632–9638.

Arnt, J. (1982). Pharmacological specificity of conditioned avoidance response Inhibition in rats: Inhibition by neuroleptics and correlation to dopamine receptor blockade. *Acta Pharmacol. Toxicol. (Copenh.), 51*(4), 321–329.

Balleine, B. W., Garner, C., Gonzalez, F., & Dickinson, A. (1995). Motivational control of heterogeneous instrumental chains. *Journal of Experimental Psychology: Animal Behaviour Processes, 21*, 203–217.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of neuroscience, 18*(1), 428–437.

Beck, A. T., & Rector, N. A. (2003). A cognitive model of hallucinations. *Cognitive Therapy and Research, 27*(1), 19–52.

Bell, D. S. (1973). The experimental reproduction of amphetamine psychosis. *Archives of General Psychiatry, 29*, 35–40.

Beninger, R. J. (1989). Dissociating the effects of altered dopaminergic function on performance and learning. *Brain Research Bulletin, 23*, 365–371.

Beninger, R. J., & Hahn, B. L. (1983). Pimozide blocks establishment but not expression of amphetamine-produced environment-specific conditioning. *Science, 220*, 1304–1306.

Beninger, R. J., Mason, S. T., Phillips, A. G., & Fibiger, H. C. (1980a). The use of conditioned suppression to evaluate the nature of neuroleptic-induced avoidance deficits. *J. Pharmacol. Exp. Ther., 213*(3), 623-627.

Beninger, R. J., Mason, S. T., Phillips, A. G., & Fibiger, H. C. (1980b). The use of extinction to investigate the nature of neuroleptic-induced avoidance deficits. *Psychopharmacology (Berl.), 69*(1), 11–18.

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews, 28*, 309–369.

Berridge, K., & Robinson, T. E. (2003). Parsing reward. *Trends in Neurosciences, 26*(9), 507–513.

Berridge, K. C., & Schulkin, J. (1989). Palatability shift of a salt-associated incentive during sodium depletion. *Quarterly Journal of Experimental Psychology, 41B*(2), 121–138.

Black, A. H. (1963). The effects of CS-US interval on avoidance conditioning in the rat. *Canadian Journal of Psychology, 17*(2), 174–182.

Blackburn, J. R., & Phillips, A. G. (1989). Blockade of acquisition of one-way conditioned avoidance responding by haloperidol and metoclopramide but not

by thioridazine or clozapine: Implications for screening new antipsychotic drugs. *Psychopharmacology, 98*, 453–459.

Cador, M., Robbins, T. W., & Everitt, B. J. (1989). Involvement of the amygdala in stimulus-reward associations: Interaction with the ventral striatum. *Neuroscience, 30*(1), 77–86.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal coretx. *Neuroscience and Biobehavioral Reviews, 26*(3), 321–352.

Cardinal, R. N., Pennicott, D. R., Sugathapala, C. L., Robbins, T. W., & Everitt, B. J. (2001). Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science, 292*(5526), 2499–2501.

Cardinal, R. N., Robbins, T. W., & Everitt, B. J. (2000). The effects of d-amphetamine, chlordiazepoxide, alpha-flupenthixol and behavioural manipulations on choice of signalled and unsignalled delayed reinforcement in rats. *Psychopharmacology, 152*, 362–375.

Catania, A. C. (in press). Attention-deficit/hyperactivity disorder (ADHD): Delay-of-reinforcement gradients and other behavioral mechanisms. *Behavioral and Brain Sciences*.

Connell, P. H. (1958). *Amphetamine psychosis*. London: Chapman and Hall.

Cook, L., & Catania, A. C. (1964). Effects of drugs on avoidance and escape behavior. *Federation Proceedings, 23*, 818–835.

Cook, L., & Weidley, E. (1957). Behavioral effects of some psychopharmacological agents. *Ann. N.Y. Acad. Sci., 66*, 740–752.

Courvoisier, S. (1956a). Pharmacodynamic basis for the use of chlorpromazine in psychiatry. *Quarterly Review of Psychiatry and Neurology, 17*(1), 25–37.

Courvoisier, S. (1956b). Pharmacodynamic basis for the use of chlorpromazine in psychiatry. *Journal of Clin. Exp. Psychophathol. Quart. Rev. Psychiat. Neurol., 17*, 25–37.

Cousins, M. S., Atherton, A., Turner, L., & Salamone, J. D. (1996). Nucleus accumbens dopamine depletions alter relative response allocation in a T-maze cost/benefit task. *Behavioural Brain Research, 74*, 189–197.

Davidson, A. B., & Weidley, E. (1976). Differential effects of neuroleptic and other psychotropic agents on acquisition of avoidance in rats. *Life Sci., 18*(11), 1279–1284.

Davis, W. M., Capehart, J., & Llewellin, W. L. (1961). Mediated acquisition of a fear-motivated response and inhibitiory effects of chlorpromazine. *Psychopharmacologia, 2*, 268–276.

Daw, N. D., Courville, A. C., & Touretzky, D. S. (2004). Timing and partial observability in the dopamine system. In S. Still, W. Bialek, & L. Botlou (Eds.), *advances in neural information processing systems, 16*. Cambridge, MA: MIT Press.

Dayan, P. (2002). Motivated reinforcement learning. In T. G. Dietterich, S. Becker, & Z. Ghahramani, (Eds.) *Advances in neural information processing systems, 14*. Cambridge, MA: MIT Press.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation and reinforcement learning. *Neuron, 36*, 285–298.

de Wit, H., Enggasser, J. L., & Richards, J. B. (2002). Acute administration of d-amphetamine decreases impulsivity in healthy volunteers. *Neuropsychopharmacology, 27*(5), 813–825.

Dews, P. B., & Morse, W. H. (1961). Behavioral pharmacology. *Annual Review of Pharmacology, 1*, 145–174.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Disssociation in prefrontal cortex of affective and attentional shifts. *Nature, 380*(6569), 69–72.

DiChiara, G. (1999). Drug addiction as dopamine-dependent associative learning disorder. *European Journal of Pharmacology, 375*, 13–30.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.

Dickinson, A. (1987). Instrumental performance following saccharin prefeeding. *Behavioural Processes, 14*, 147–154.

Dickinson, A., Nicholas, D. J., & Adams, C. D. (1983). The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *Quarterly Journal of Experimental Psychology, 35B*, 35–51.

Dickinson, A., Smith, J., & Mirenowicz, J. (2000). Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. *Behavioral Neuroscience, 40*, 468–483.

Ettenberg, A. (1989). Dopamine, neuroleptics and reinforced behavior. *Neuroscience and Biobehavioral Reviews, 13*, 105–111.

Ettenberg, A., Koob, G. F., & Bloom, F. E. (1981). Response artifact in the measurement of neuroleptic-induced anhedonia. *Science, 213*, 357–359.

Evenden, J. L., & Robbins, T. W. (1983). Dissociable effects of d-amphetamine, chlordiazepoxide and alpha-flupenthixol on choice and rate measures of reinforcement in the rat. *Psychopharmacology, 79*, 180–186.

Everitt, B. J., Morris, K. A., O'Brien, A., & Robbins, T. W. (1991). The basolateral amygdala-ventral striatal system and conditioned place preference: Further evidence of limbic-striatal interactions underlying reward-related processes. *Neuroscience, 42*, 1–18.

Everitt, B. J., & Robbins, M. C. T. W. (1989). Interactions between the amygdala and ventral striatum in the stimulus-reward associations: Studies using a second-order schedule of sexual reinforcement. *Neuroscience, 30*(1), 63–75.

Fibiger, H. C., Carter, D. A., & Phillips, A. G. (1976). Decreased intracranial self-stimulation after neuroleptics of 6-hydroxydopamine: Evidence for mediation by motor deficits rather than by reduced reward. *Psychopharmacology, 47*, 21–27.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science, 299*, 1898–1902.

Fowler, S. C., LaCerra, M. M., & Ettenberg, A. (1986). Effects of haloperidol on the biophysical characteristics of operant responding: Implications for motor and reinforcement processes. *Pharmacology, Biochemistry and Behaviour, 25*, 791–796.

Frances, A., (Ed.). (2000). *Diagnostic and statistical manual of mental disorders*. Washington, DC: American Psychiatric Association.

Giordano, L. A., Bickel, W. K., Loewenstein, G., Jacobs, E. A., Marsch, L., & Badger, G. J. (2002). Mild opioid deprivation increases the degree that opioid-

dependent outpatients discount delayed heroin and money. *Psychopharmacology, 163*, 174–182.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22*, 10829–10837.

Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience, 41*(1), 1–24.

Grace, A. A. (2000). Gating of information flow within the limbic system and the pathophysiology of schizophrenia. *Brain Research Reviews, 31*, 330–341.

Grilly, D. M., Johnson, S. K., Minardo, R., Jacoby, D., & LaRiccia, J. (1984). How do tranquilizing agents selectively inhibit conditioned avoidance responding? *Psychopharmacology, 84*, 262–267.

Hollerman, J., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1*, 304–309.

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience, 96*(4), 651–656.

Horvitz, J. C. (2002). Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behavioural Brain Research, 137*, 65–74.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 249–270). Cambridge, MA: MIT press.

Hunt, H. F. (1956). Some effects of drugs on classical (type S) conditioning. *Ann. N.Y. Acad. Sci., 65*, 258–267.

Ikemoto, S., & Panksepp, J. (1999). The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Research Reviews, 31*(1), 6–41.

Irwin, S. (1958). Factors influencing acquisition of avoidance behaviour and sensitivity to drugs. *Fed. Proc., 17*, 380.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Experimental Brain Research, 11*, 376–386.

Janssen, P. A. J., Niemegeers, C. J. E., & Schellekens, K. H. L. (1965). Is it possible to predict the clinical effects of neuroleptic drugs (major tranquilizers) from animal data? *Arzneimittelforschung, 15*, 104–117.

Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., & Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron, 40*, 1251–1257.

Joseph, M. H., Datla, K., & Young, A. M. J. (2003). The interpretation of the measurement of nucleus accumbens dopamine by vivo dialysis: The kick, the craving or the cognition. *Neuroscience and Biobehavioral Reviews, 27*, 527–541.

Kaelbling, L., Littman, M., & Moore, A. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence, 4*, 237–285.

Kamin, L. J. (1954). Traumatic avoidance learning: The effects of CS-US interval with a trace-conditioning procedure. *Journal of Comparative and Physiological Psychology, 47*, 65–72.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1991). *Principles of neural science*. New York: Elsevier Science.

Kauer, J. A. (2003). Addictive drugs and stress trigger a common change at VTA synapses. *Neuron, 37*, 549–550.

Key, B. J. (1961). The effects of drugs on discrimination and sensory generalisation of auditory stimuli in cats. *Psychopharmacologia, 2*, 352–363.

Kilts, C. D. (2001). The changing roles and targets for animal models of schizophrenia. *Biological Psychiatry, 50*(11), 845–855.

Low, L. A., Eliasson, M., & Kornetsky, C. (1966). Effects of chlorpromazine on avoidance acquisition as a function of CS-US interval length. *Psychopharmacologia, 10*, 148–154.

Low, L. A., & Low, H. L. (1962). Effects of CS-US interval upon avoidance responding. *Journal of Comparative and Physiological Psychology, 55*(6), 1059–1061.

Maffii, G. (1959). The secondary conditioned response of rats and effects of some psychopharmacological agents. *Journal of Pharmacy and Pharmacology, 11*, 129–139.

Maher, B., & Ross, J. S. (1984). Delusions. In H. E. Adams, & P. Sutker (Eds.), *Comprehensive handbook of psychopathology* (pp. 383-409). New York: Plenum Press.

Masterman, D. L., & Cummings, J. L. (1997). Frontal-subcortical circuits: The anatomical basis of executive, social and motivated behaviors. *Journal of Psychopharmacology, 11*(2), 107–114.

McClure, S. M., Daw, N., & Montague, P. R. (2003). A computational substrate for incentive salience. *Trends in Neuroscience, 26*(8), 423–428.

Miller, R. E., Murphy, J. V., & Mirsky, A. (1957). The effect of chlorpromazine on fear-motivated behavior in rats. *Journal of Pharmacol. and Exper. Therap., 120*, 379–387.

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron, 36*, 265–284.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature, 377*, 725–728.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16*(5), 1936–1947.

Moore, H., West, A. R., & Grace, A. A. (1999). The regulation of forebrain dopamine transmission: Relevance to the pathophysiology and psychopathology of schizophrenia. *Biological Psychiatry, 46*, 40–55.

Morpurgo, C. (1965). Drug-induced modifications of discriminated avoidance behavior in rats. *Psychopharmacologia, 8*, 90–99.

Nader, K., & LeDoux, J. (1999). The dopaminergic modulation of fear: Quinpirole impairs the recall of emotional memories in rats. *Behavioral Neuroscience, 113*(1), 152–165.

Nishijo, H., Ono, T., & Nishino, H. (1988). Single neuron responses in amygdala of alert monkey during complex senosry stimulation with affective significance. *Journal of Neuroscience, 8*, 3570–3583.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 28*, 329–337.

O'Doherty, J. P., Deichmann, R., Critchley, H. D., & Dolan, R. J. (2002). Neural responses during anticipation of primary taste reward. *Neuron, 33*, 815–826.

Ogren, S. O., & Archer, T. (1994). Effects of typical and atypical antipsychotic drugs on two-way active avoidance: Relationship to DA receptor blocking profile. *Psychopharmacology (Berl.), 114*(3), 383–391.

Pennartz, C. M. A., Groenewegen, H. J., & Silva, F. H. L. D. (1994). The nucleus accumbens as a complex of functionally distinct Neuronal ensembles: An integration of behavioural, electrophysiological and anatomical data. *Progress in Neurobiology, 42*, 719–761.

Phares, V. (2003). *Understanding abnormal child psychology*. New York: Wiley.

Phillips, P. E. M., Stuber, G. D., Helen, M. L. A. V., Wightman, R. M., & Carelli, R. M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature, 422*, 614–618.

Ponsluns, D. (1962). An analysis of chlorpromazine-induced suppression of the avoidance response. *Psychopharmacologia, 3*, 361–373.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences, 22*(4), 146–151.

Richards, J. B., Sabol, K. E., & de Wit, H. (1999). Effects of methamphetamine on the adjusting amount of procedure, a model of impulsive behavior in rats. *Psychopharmacology, 146*, 432–439.

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology, 81*(1), 1–11.

Robbins, T. W., Cador, M., Taylor, J. R., & Everitt, B. J. (1989). Limbic-striatal interactions in reward-related processes. *Neuroscience and Biobehavioural Reviews, 13*, 155–162.

Rolls, E. T., Rolls, B. J., Kelly, P. H., Shaw, S. G., Wood, R. J., & Dale, R. (1974). The relative attenuation of self-stimulation, eating and drinking produced by dopamine-receptor blockade. *Psychopharmacology, 38*, 219–230.

Salamone, J. D., Cousins, M. S., & Bucher, S. (1994). Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behavioural Brain Research, 65*, 221–229.

Salamone, J. D., Cousins, M. S., & Snyder, B. J. (1997). Behavioural functions of nucleus accumbens dopamine: Empirical and conceptual problems with the anhedonia hypothesis. *Neuroscience and Biobehavioural Reviews, 21*(3), 341–359.

Salamone, J. D., Kurth, P. A., McCullough, L. D., Sokolowski, J. D., & Cousins, M. S. (1993). The role of brain dopamine in response initiations: Effects of

haloperidol and regionally-specific dopamine depletions on the local rate of instrumental responding. *Brain Research, 628*, 218–226.

Salamone, J. D., Steinpreis, R. E., McCullough, L. D., Smith, P., Grebel, D., & Mahan, K. (1991). Haloperidol and nucleus accumbens dopamine depletion suppress lever pressing for food but increase free food consumption in a novel food-choice procedure. *Psychopharmacology, 104*, 515–521.

Salamone, J. D., Wisniecki, A., Carlson, B. B., & Correa, M. (2001). Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience, 105*(4), 863–870.

Schmajuk, N. A. (1988). The hippocampus and the classically conditioned nictitating membrane response: A real-time attentional-associative model. *Psychobiology, 16*(1), 20–35.

Schmajuk, N. A., Cox, L., & Gray, J. A. (2001). Nucleus accumbens, entorhinal cortex and latent inhibition: A neural network model. *Behavioural Brain Research, 118*, 123–141.

Schneider, J. S., Sun, Z. Q., & Roeltgen, D. P. (1994). Effects of dopamine agonists on delayed response performance in chronic low-dose MPTP-treated monkeys. *Pharmacology, Biochemistry and Behaviour, 48*(1), 235–240.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593–1599.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex, 10*, 272–283.

Seeman, P., & Madras, B. K. (1998). Anti-hyperactivity medication: Methylphenidate and amphetamine. *Molecular Psychiatry, 3*(5), 386–396.

Smith, A., Li, M., Becker, S., & Kapur, S. (2004). A model of antipsychotic action in conditioned avoidance: A computational approach. *Neuropsychopharmacology, 29*(6), 1040–1049.

Solanto, M. V., Abikoff, H., Sonuga-Barke, E., Schachar, R., Logan, G. D., Wigal, T., Hechtman, L., Hinshaw, S., & Turkel, E. (2001). The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: A supplement to the NIMH multimodal treatment study of AD/HD. *Journal of Abnormal Child Psychology, 29*(3), 215–228.

Stark, H., Bischof, A., & Scheich, H. (1999). Increase of extracellular dopamine in prefrontal cortex of gerbils during acquisition of the avoidance strategy in the shuttle box. *Neuroscience Letters, 264*, 77–80.

Suri, R. E. (2001). Anticipatory responses of dopamine neurons and cortical neurons reproduced by internal model. *Experimental Brain Research, 140*, 234–240.

Suri, R. E. (2002). *TD* models of reward predictive responses in dopamine neurons. *Neural Networks: Special Issue on Computational Models of Neuromodulation, 15*(4-6), 523–533.

Suri, R. E., Bargas, J., & Arbib, M. A. (2001). Modeling functions of striatal dopamine modulation in learning and planning. *Neuroscience, 103*(1), 65–85.

Suri, R., & Schultz, W. (1998). Learning of sequential movements by neural network model with dopamine-like reinforcement signal. *Experimental Brain Research, 121*, 350–354.

Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine Learning, 3*, 9–44.

Sutton, R. S., & Barto, A. G. (1981). An adaptive network that constructs and uses an internal model of its world. *Cognition and Brain Theory, 4*(3), 217–246.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.

Talk, A. C., Gandhi, C. C., & Matzel, L. D. (2002). Hippocampal function during behaviourally silent association learning: Dissociation of memory storage and expression. *Hippocampus, 12*, 648–656.

Taylor, J. R., & Robbins, T. W. (1984). Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. *Psychopharmacology, 84*, 405–412.

Tremblay, L., & Schultz, W. (2000a). Modifications of reward expectation–related neuronal activity during learning in primate orbitofrontal cortex. *Journal of Neurophysiology, 83*, 1877–1885.

Tremblay, L., & Schultz, W. (2000b). Reward-related neuronal activity during go–nogo task performance in primate orbitofrontal cortex. *Journal of Neurophysiology, 83*, 1864–1876.

van der Heyden, J. A. M., & Bradford, L. D. (1988). A rapidly acquired one-way conditioned avoidance procedure in rats as a primary screening test for antipsychotics: Influence of shock intensity on avoidance performance. *Behavioural Brain Research, 31*, 61–67.

Wade, T. R., de Wit, H., & Richards, J. B. (2000). Effects of dopaminergic drugs on delayed reward as a measure of impulsive behavior in rats. *Psychopharmacology, 150*, 90–101.

Wadenberg, M. G., Soliman, A., Vanderspek, S. C., & Kapur, S. (2001). Dopamine D2 receptor occupancy is a common mechanism underlying animal models of antipsychotics and their clinical effects. *Neuropsychopharmacology, 25*(25), 633–641.

Wadenberg, M. L., & Hicks, P. B. (1999). The conditioned avoidance response test reevaluated: Is it a sensitive test for the detection of potentially atypical antipsychotics? *Neurosci. Biobehav. Rev., 23*(6), 851–862.

Wadenberg, M.-L. G., Kapur, S., Soliman, A., Jones, C., & Vaccarino, F. (2000). Dopamine D2 receptor occupancy predicts catalepsy and the supression of conditioned avoidance response behaviour in rats. *Psychopharmacology, 150*, 422–429.

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412*, 43–48.

Wilkinson, L. S., Humby, T., Killcross, A. S., Torres, E. M., Everitt, B. J., & Robbins, T. W. (1998). Dissociations in dopamine release in medial prefrontal cortex and ventral striatum during the acquisition and extinction of classical aversive conditioning in the rat. *European Journal of Neuroscience, 10*, 1019–1026.

Wilson, M., & Daly, M. (2004). Do pretty women inspire men to discount the future? *Biology letters, 271*(S4), 177–179.

Wise, R. A. (1982). Neuroleptics and operant behavior: The anhedonia hypothesis. *Behavioural and Brain Sciences, 5*, 39–87.

Wise, R. A. (2002). Brain reward circuitry: Insights from unsensed incentives. *Neuron, 36*, 229–240.

Wise, R. A., & Schwartz, H. V. (1981). Pimozide attenuates acquisition of lever-pressing for food in rats. *Pharmacology, Biochemistry and Behavior, 15*, 655–656.

Wise, R. A., Spindler, J., DeWit, H., & Gerber, G. J. (1978). Neuroleptic-induced "anhedonia" in rats: Pimozide blocks reward quality of food. *Science, 201*, 262–264.

Young, A. M. J., Ahier, R. G., Upton, R. L., Joseph, M. H., & Gray, J. A. (1998). Increased extracellular dopamine in the nucleus accumbens of the rat during associative learning of neutral stimuli. *Neuroscience, 83*(4), 1175–1183.

Received January 9, 2004; accepted July 2, 2004.